

Data processing method applying Principal Component Analysis and Spectral Angle Mapper for imaging spectroscopic sensors

P.B. Garcia-Allende, O.M. Conde, J. Mirapeix, A.M. Cubillas, J.M. Lopez-Higuera
Photonics Engineering Group, University of Cantabria, Avda. de los Castros S/N, 39005 Santander,
Spain

ABSTRACT

A data processing method for hyperspectral images is presented. Each image contains the whole diffuse reflectance spectra of the analyzed material for all the spatial positions along a specific line of vision. This data processing method is composed of two blocks: data compression and classification unit. Data compression is performed by means of Principal Component Analysis (PCA) and the spectral interpretation algorithm for classification is the Spectral Angle Mapper (SAM). This strategy of classification applying PCA and SAM has been successfully tested on the raw material on-line characterization in the tobacco industry. In this application case the desired raw material (tobacco leaves) should be discriminated from other unwanted spurious materials, such as plastic, cardboard, leather, candy paper, etc. Hyperspectral images are recorded by a spectroscopic sensor consisting of a monochromatic camera and a passive Prism-Grating-Prism device. Performance results are compared with a spectral interpretation algorithm based on Artificial Neural Networks (ANN).

Keywords: Principal Component Analysis (PCA), Spectral Angle Mapper (SAM), absorption spectroscopy, hyperspectral images, raw material on-line characterization, pattern recognition, classification analysis.

1. INTRODUCTION

Spectroscopy measures the radiant intensity and energy of the interaction between light and any material to determine its molecular and dynamic structure. In absorption spectroscopy the compound that interacts with light behaves as a passive element. It absorbs some of the emitted photons depending on their wavelength, “*spectral signature*”. Light which is not absorbed can be transmitted through the sample of the compound or diffusely reflected in it. The spectrum of this reflected light, also known as diffuse reflectance, is measured. Afterwards, this spectrum is processed and the identification, classification or discrimination of the material or its chemical components is performed.

Imaging spectroscopy takes advance and is based on this technique, where the spectrum of diffuse reflectance in all the points across a spatial line is measured at the same time by means of an imaging spectrometer. Potential industrial applications of this hyperspectral technique are increasing [1-3] with clear examples for raw material discrimination and classification at the input chains of manufacturing processes.

For the efficient application of imaging spectroscopy to the on-line monitoring of industrial processes real time analysis conditions must be satisfied. In the particular case of raw material characterization, unwanted spurious materials have to be rejected as they are transported intermingled with the wanted raw material on the conveyor belt in the production plant. This paper presents the performance of the Spectral Angle Mapper (SAM) as the spectral interpretation algorithm and compares it with Artificial Neural Networks employed in a previous work [4]. SAM has been selected because its linearity makes it simple and temporarily efficient. Principal Component Analysis is used prior to the classification to attain a representative spectral fingerprint, compressing the large amount of data generated in the processed due to the employment of an imaging spectroscopy technique.

2. SPECTRAL SOFTWARE PROCESSING

As aforementioned the implemented spectral interpretation process has two stages: PCA [5] is first used to perform dimensionality reduction and redundancy elimination. Secondly, SAM [6] has been adopted for the determination of both presence and position of the unwanted spurious materials. A block-diagram of the proposed spectral analysis is depicted in Figure 1.

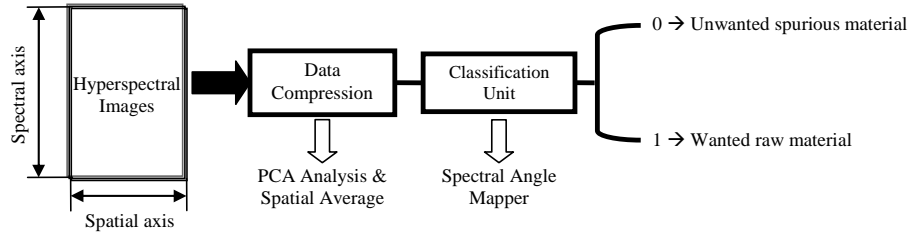


Fig. 1. Spectral analysis processing scheme.

2.1. Principal Component Analysis

Data compression carries out the reduction of both spectral and spatial axes of the captured hyperspectral image. The spatial axis is simply averaged. This can be done because the system of mechanical extraction of unwanted material has low spatial resolution. This processing also helps in the reduction of noise in the measurement.

The treatment of the spectral axis is more sophisticated. It is based on PCA and it is accomplished by the projection of the captured spectra of each spatial point over a different vectorial basis. The new basis vectors are those directions of the initial data containing the most relevant information. After mean subtraction, the covariance matrix of the data is computed and their eigenvalues and eigenvectors are obtained. In the designed algorithm, those eigenvalues smaller than a thousandth of the maximum eigenvalue are ignored. In this way, the captured spectra (from 400 to 1000 nm), which initially lies in a 640-dimensional space, is reduced to 18 wavelengths, providing, therefore, a 99% data compression rate. This reduction allows a great enhancement in the computational performance, which minimizes the classification time, without causing an appreciable increase in the classification error.

2.2. Spectral Angle Mapper

Spectral Angle Mapper is a simple algorithm based on the measurement of the spectral similarity between two spectra. This spectral similarity, α , is obtained considering each spectrum as a vector in a n -dimensional space [6]:

$$\alpha = \cos^{-1} \left(\frac{\sum_{i=1}^n s_1 s_2}{\left(\sum_{i=1}^n s_1^2 \right)^{1/2} \cdot \left(\sum_{i=1}^n s_2^2 \right)^{1/2}} \right) \quad (1)$$

where n is the number of spectral bands, and s_1 and s_2 are the two spectra. Small angles between the two spectra indicate high similarity and high angles indicate low similarity. As PCA has been firstly applied, the dimensional space of the spectral axis contains now only 18 bands.

For the proper implementation of the classification method based on SAM, two stages should be followed:

1. First, in the *training stage*, reference spectrum of the wanted and the unwanted material should be collected. For the wanted material (tobacco leaves) a reference spectrum has been calculated by averaging 30 *homogeneous* (the same kind of material in all the points of the line of vision) images (2880 spectra). The unwanted material reference spectrum has been determined by averaging 24 images (2304 spectra) of the typical spurious materials in this application (foil, leather, plastics, etc.).
2. In the *test stage*, a new unknown image is classified. The spectral similarities α_{t-w} (between the spectrum of each spatial point, s_t , of the test image and the wanted reference spectrum) and α_{t-uw} (between the test spectrum and the unwanted reference spectrum) are calculated. If $\alpha_{t-w} < \alpha_{t-uw}$ the test spectrum, s_t , is closer to the wanted

reference spectrum and, therefore, that position contains tobacco. On the other hand, $\alpha_{t-uw} < \alpha_{t-w}$ indicates that s_t corresponds to unwanted material. A schematic of the classification procedure is depicted in Figure 2, using a dimensionality of $n=3$ for visualization purposes.

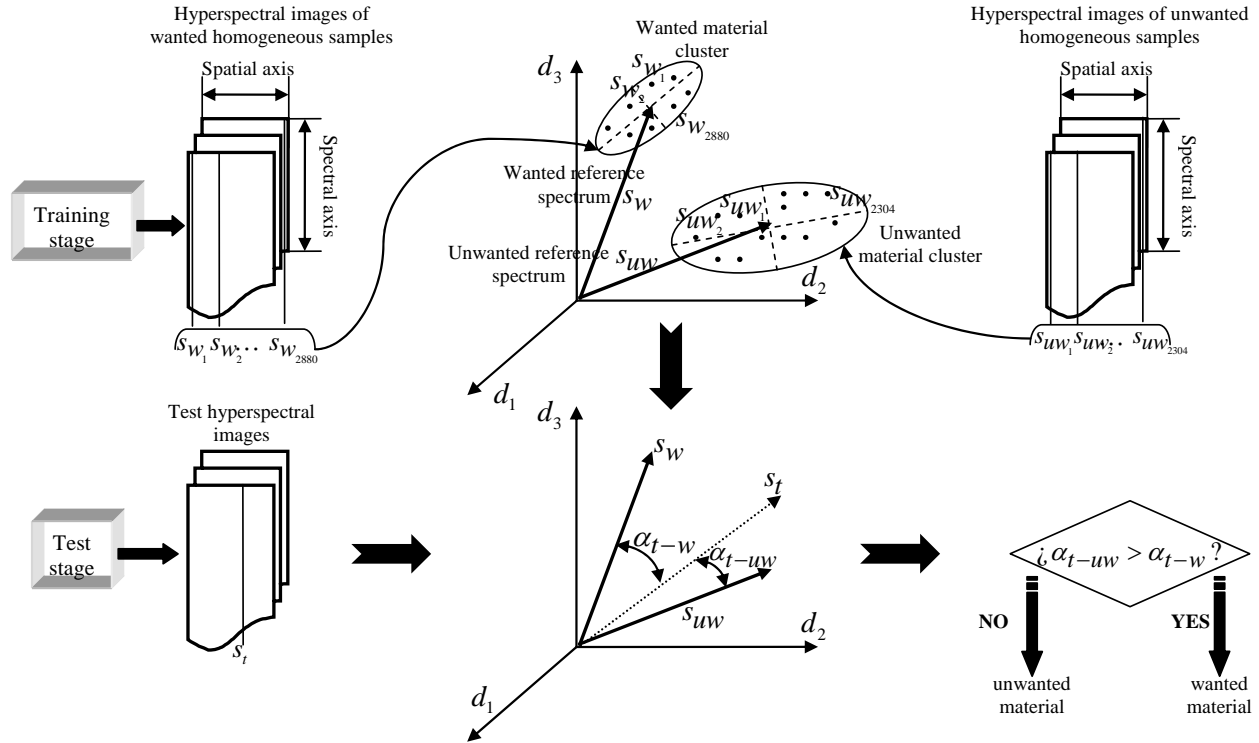


Fig. 2. Schematic description of the spectral interpretation algorithm.

3. EXPERIMENTAL ISSUES

A block diagram of the experimental setup is shown in Figure 3. The acquisition system includes: an illumination system consisting of two halogen floodlights with a power rating of 500W, front objective lens Zoom Navitar 7000, a commercial PGP (Prism-Grating-Prism) imaging spectrograph [1] known as ImSpector and the monochrome digital camera Pixelink PL-A741. A computer equipped with IEEE1394 interface for image data acquisition employed to perform data compression and the discrimination of the wanted from the unwanted raw material.

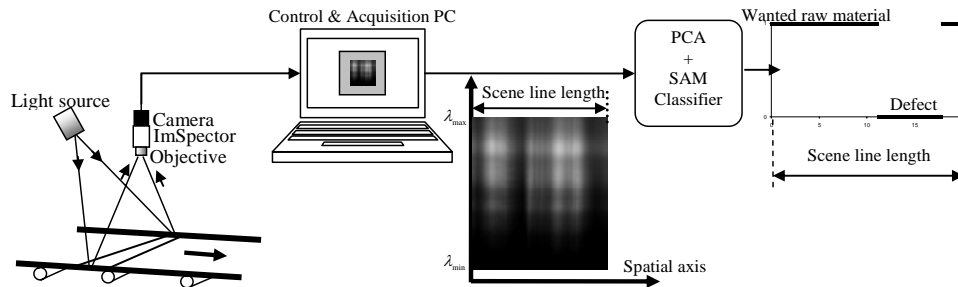


Fig. 3. Block diagram of the entire experimental system.

Test samples for validation purposes are presented in Figure 4. In Figure 4 (a) tobacco leaf blends are mixed with foil. From $x \approx 11$ mm to $x \approx 18$ mm the output of the system shows the clear detection of foil as unwanted material (Figure 4 (b)). Figure 4 (c) contains a piece of leather on a bunch of tobacco leaves. Figure 4 (d) shows the SAM outputs, where an unwanted material section from $x \approx 11$ mm to $x \approx 16$ mm is labeled. A “false alarm”, a wanted spectrum that is classified as unwanted, occurs in the interface between the piece of leather and the final tobacco leaves; this has been highlighted with a circle in the graph. Obviously, this fact increases the classification error but, from the industrial point of view, it is

admissible compared to the confusion of an unwanted spectrum with tobacco. Finally, a sample consisting of tobacco leave blends and a brown leaf of a vegetable different from tobacco is presented in Figure 4 (e), and their associated classification system outputs in Figure 4 (f). The brown leaf is correctly classified from $x \approx 2$ mm to $x \approx 12$ mm, showing that the proposed technique is not based on a colorimetric analysis.

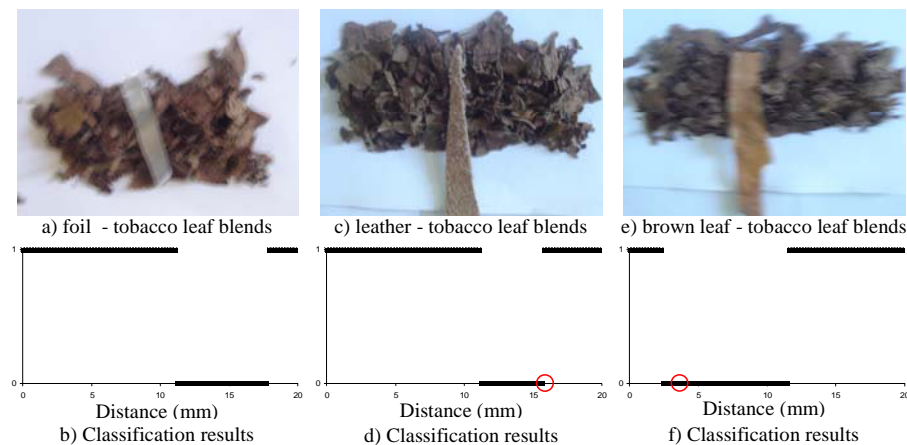


Fig. 4. Test samples and their associated SAM results.

A total of 32 hyperspectral images (3072 spectra) were used as test data achieving a classification accuracy of 90.54%. Although with the technique based on ANN [4] a 99.12% classification accuracy was achieved, the main power of the SAM algorithm is that each image is classified in a mean time of ~ 4 ms implying an extremely high ($\sim 99\%$) reduction in the execution time compared to the previous technique [4]. All these comparisons have been carried out in a Pentium (R) 4 processor of 2GHz with 1.00 GB de DDR-RAM of 2.01 GHz.

4. CONCLUSIONS

A specifically designed data processing method for hyperspectral images consisting of Principal Component Analysis and Spectral Angle Mapper has been successfully checked in a raw-material on-line characterization system. The use of PCA as a pre-processing stage improves the computational performance of the SAM. Several test samples have been reported, demonstrating that, the SAM efficiently discriminates different defects. The key point of this algorithm is its efficient discrimination due to the simplicity and linearity of the SAM algorithm.

ACKNOWLEDGEMENTS

This work has been co-supported by the Spanish TEC'2004-05936-C02-02 and TEC'2005-08218-C02-02 projects.

REFERENCES

1. T. Vaarala, M. Aikio and H. Keraenen, An advanced prism-grating-prism imaging spectrograph in on-line industrial applications, *Proceedings of the SPIE*, Volume 3101 (1997), pp. 322-30.
2. E. Herrala, T. Hyvarinen, O. Voutilainen and J. Lammasniemi, An optoelectronic sensor system for industrial multipoint and imaging spectrometry, *Sensors and Actuators A (Physical)*, Volume A61, n 1-3 (1997), pp. 335-338.
3. D.P. Ariana, R. Lu, D.E. Guyer, Near-infrared hyperspectral reflectance imaging for detection of bruises on pickling cucumbers, *Computers and Electronics in Agriculture*, Volume 53, n 1 (2006), pp. 60-70
4. P.B. García Allende, O.M. Conde, A.M. Cubillas, C. Jáuregui and J.M. López-Higuera, New Raw Material Discrimination System Based on a Spatial Optical Spectroscopy Technique, *Sens. Actuators A: Phys. In Print*.
5. J. Workman Jr, A.W. Springsteen, *Applied Spectroscopy: A compact reference for Practitioners*, first ed., Academic Press Limited, London, 1998.
6. G. Girouard, A. Bannari, A. El Harti, A. Desrochers, Validated Spectral Angle Mapper Algorithm for Geological Mapping: Comparative Study between Quickbird and Landsat-TM, *Geo-Imagery Bridging Continents*, Istanbul, July 12–23, 2004, pp. 599–604.